Nucleic Acids Research, 1993, Vol. 21, No. 12 2861–2865

Bacteriophage Nf DNA region controlling late transcription: structural and functional homology with bacteriophage φ29

Beatriz Nuez+ and Margarita Salas*
Centro de Biología Molecular (CSIC-UAM), Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain

Received March 16, 1993; Revised and Accepted May 12, 1993 EMBL accession no. X72380

ABSTRACT

The putative region for the control of late transcription of the *Bacillus subtilis* phage Nf has been identified by DNA sequence homology with the equivalent region of the evolutionary related phage φ29. A similar arrangement of early and late promoters has been detected in the two phages, suggesting that viral transcription could be regulated in a similar way at late times of the infection. Transcription of late genes requires the presence of a viral early protein, gpF in phage Nf and p4 in phage φ29, being the latter known to bind to a DNA region located upstream from the phage φ29 late promoter. We have identified a DNA region located upstream from the putative late promoter of phage Nf that is probably involved in binding protein gpF. Furthermore, we show that the phage φ29 protein p4 is able to bind to this region and activate transcription from the phage Nf putative late promoter. Sequence alignment has also revealed the existence of significant internal homology between the two early promoters contained in this region of each phage.

INTRODUCTION

Bacteriophage φ29 belongs to a group of phages that infect *B.subtilis* and are characterized by having a double-stranded linear DNTA genome that is replicated by a protein-priming initiation process (1). Different criteria allowed their classification in at least three families that include phages φ29, φ15, PZA and PZE (group A), phages Nf, M2 and B103 (group B), and phage GA1 (group C). The first two groups are more closely related to each other than to group C (1, 2). Phages Nf and φ29 have genomes of slightly different size and their physical map is not conserved (2), however, the arrangement of their early and late genes (3, 4) is homologous (Figure 1), suggesting that they might have a similar transcriptional map. Late transcription of these phages requires the expression of an early viral protein: protein p4 in phage φ29 (5) and protein gpF in phage Nf (6), known to be 70% identical (7). Most of phage φ29 genes are transcribed from a group of promoters located in a small DNA region (8). Early genes, mainly contained in the left genome end, are transcribed from the early PA2c and PA2b promoters (8), whereas all late genes are transcribed from the divergent PA3 promoter (5). The PA3 promoter is only efficiently transcribed when protein p4 binds to a DNA region located between positions −58 and −104 relative to the transcription initiation start point (9, numbering according to 10). This region includes the inverted repeat 5′-AACTTTTT-15 base pairs (bp)-AAAATGTT-3′ that is recognized by protein p4, and that overlaps with the −35 consensus region of the PA2b early promoter (9, 11). This promoter distribution determines that binding of protein p4 activates transcription of late genes from the PA3 promoter, while represses transcription of early genes from the PA2b promoter (12), which is the main early promoter *in vivo* (13). Therefore, phage φ29 late transcription is tightly regulated by protein p4.

In this report, we have sequenced the phage Nf DNA region equivalent to the one that controls late transcription in phage φ29. In this region of phage Nf, we have found putative promoters homologous to the PA2b, PA2c and PA3 promoters of phage φ29, and a region likely involved in binding the phage transcriptional regulator, protein gpF. Transcriptional analysis of this phage Nf region has been done in two different salt conditions and in the absence or presence of the phage φ29 transcriptional regulator, protein p4. The results obtained suggest that the promoters identified in phage Nf probably participate in late transcription regulation in a similar way that the PA2b, PA2c and PA3 promoters in phage φ29. Sequence alignment of the early promoters has revealed the existence of internal homology between them. Additionaly, comparison of the putative gpF and the protein p4 binding sites has provided valuable information that might help to understand the way in which these proteins bind to DNA and activate transcription.

MATERIALS AND METHODS

DNA manipulations

Plasmid DNA was prepared and manipulated following standard methods (14); phage DNA was isolated as described (15). All enzymatic reactions were done according to the supplier except where otherwise is stated. Sequencing reactions were done with

* To whom correspondence should be addressed

+ Present address: National Institute for Medical Research (MRC) The Ridgeway, Mill Hill, London NW7 1AA, UK

Sequenase version 2.0 from United States Biochem. Corp. DNA restriction fragments were selectively labelled at one of their 3' ends with the Klenow fragment of *Escherichia coli* DNA polymerase I. Chemical sequencing reactions were performed as described (16). The phage Nf DNA region with the PNfA3 promoter and the putative gpF binding site was cloned into the *Sma*I site of pUC18Ω (17) producing the pNfA3 plasmid, by introducing a 607 bp long *Dra*I fragment that contained 63 bp of the internal *Hind*III fragment and 544 bp of the 626 bp *Hind*III–*Eco*RI fragment (Figure 1C).

In vitro transcription assays

Transcription reactions were done in the absence or presence of 1 μg of protein p4 with 100 ng of either the phage Nf DNA *Hinf*I–*Sau*3A fragment, 794 bp long, or the phage φ29 DNA *Hind*III H fragment, 759 bp long. Reactions were done in a final volume of 25 μl, in 25 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 20 mM NaCl, 200 μM each ATP, CTP and GTP, 80 μM [α-^{32}P]UTP (2μCi), and 4% glycerol. Ammonium sulphate concentration was either 92 mM or 18.5 mM, considered as high or low ionic strength conditions, respectively. Transcription was started by addition of 0.25 μg of *B.subtilis* σ^ARNA polymerase, incubated for 10 min at 37 °C, and stopped with 0.15% SDS, 0.15% tRNA. Samples were filtered through Sephadex G50 spun columns, ethanol precipitated, and analyzed in denaturing polyacrylamide gels.

DNase I footprinting

The *Xba*I–*Kpn*I fragment used, 650 bp long, was purified from pNfA3 and was labelled at the coding strand for early genes. Protein–DNA complexes were formed in a final volume of 20 μl, with 1 μg of protein p4 and/or 0.5 μg of σ^ARNA polymerase, in 25 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 25 mM NaCl, 15 mM ammonium sulphate, 5% glycerol, 0.1 mg/ml bovine serum albumine and 0.1 mg/ml of poly(dI-dC) as competitor DNA. Samples were incubated for 10 min at 37 °C before treatment with 1 μl (1–50 ng) of DNase I for two min at room temperature. The reaction was stopped with 2.5 μl of 0.25 M EDTA, ethanol precipitated, and the digestion products were analyzed in denaturing polyacrylamide gels.

RESULTS

Cloning and sequencing of the putative region that controls late transcription in phage Nf

The comparison of the phage Nf genetic map with that of phage φ29 (Figure 1A), suggested that promoters responsible for the transcription of the early genes located at the left genome end and of late genes, might be included within the *Eco*RI D restriction fragment of phage Nf (Figure 1B). Moreover, the *Eco*RI D fragment contained an intrinsic curvature (M.J.Otero, B.Nuez and M.Salas, unpublished results) that could be equivalent to the curvature detected in restriction fragments containing the late promoter of bacteriophage φ29 (9). The *Eco*RI D restriction fragment was purified from phage Nf DNA and digested with *Hind*III producing three smaller fragments of 626 bp, 492 bp and 447 bp (Figure 1C). The two *Eco*RI–*Hind*III external fragments, 626 bp and 447 bp long, were cloned between the *Eco*RI and *Hind*III sites of the plasmid vector pUC18Ω (17). Since no clones containing the internal *Hind*III fragment were obtained, it was directly sequenced on the *Eco*RI D fragment using specific oligonucleotides that hybridized close to the *Hind*III site of the *Eco*RI–*Hind*III fragments. The sequence obtained revealed that the phage Nf *Eco*RI D fragment contains a region homologous to the one that controls late transcription in phage φ29. As shown in Figure 2, we have identified promoters that could be homologous to the early PA2c and PA2b promoters, and to the late PA3 promoter of phage φ29, that have been named PNfA2c, PNfA2b and PNfA3, respectively. The putative late promoter of phage Nf, PNfA3, resembled that of phage φ29 in having a perfect −10 consensus sequence (TATAAT), lacking a −35 consensus sequence (TTGACA) and containing several A- or T-tracts that could be responsible of the intrinsic curvature of the fragment. Regions containing the transcription initiation point were also significantly conserved, therefore, the position homologous to the PA3 promoter +1 position in phage φ29 (10),

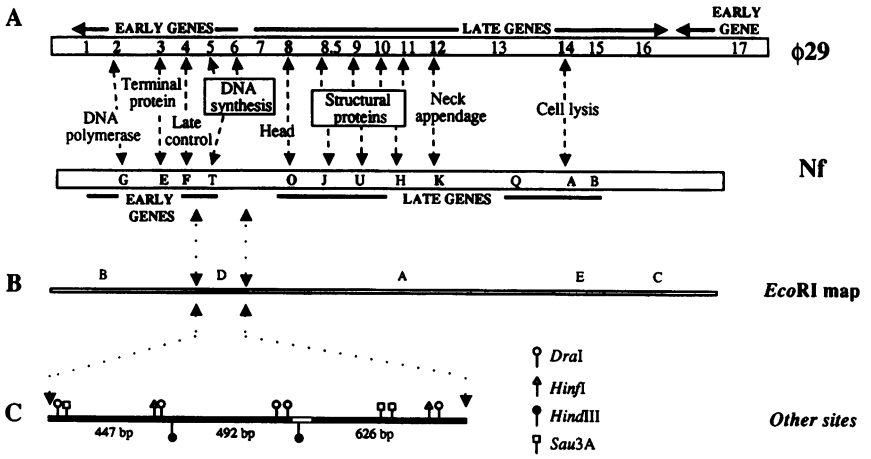

Figure 1. A) Comparison of the genetic map of phage φ29 with the genetic map of phage Nf (7), in which early and late genes are shown. Homologous genes in both phages are indicated between the two genetic maps. The transcription map of phage φ29 is defined by horizontal arrows. **B)** *Eco*RI restriction map of the Nf genome (2). **C)** *Dra*I, *Hind*III, *Hinf*I and *Sau*3A sites in the *Eco*RI D fragment. The size of the three *Hind*III fragments has been indicated. The putative gpF binding site is shown in white.

will be considered the PNfA3 promoter initiation start site in phage Nf. A putative gpF binding site could be also found located at the appropriate distance to activate transcription from PNfA3. However, this site differed from that of protein p4 in phage ϕ29: the spacing between the distal and proximal recognition sequences was one base pair longer, and the distal site contained an A- instead of a T-tract (Figure 2). Sequence alignment of the early PA2b, PA2c, PNfA2b and PNfA2c promoters revealed, in addition, the existence of an internal homology among them (Figure 3), not found with any other of the phage ϕ29 promoters (not shown). This homology included, besides the expected −10 and −35 regions, the spacer region and up to position −70, therefore containing the positions corresponding to the transcriptional regulator binding site.

In vitro activity of the phage Nf PNfA2c, PNfA2b and PNfA3 promoters

Run-off experiments were performed to analyze the *in vitro* activity of the phage Nf *Hin*fI−*Sau*3A fragment, and that of the phage ϕ29 *Hind*III H fragment, that contained, respectively, the PNfA2c, PNfA2b and PNfA3 or the PA2c, PA2b and PA3 promoters. Since proteins p4 and gpF are 70% identical (7), we also tested whether protein p4 might bind to the putative gpF site in phage Nf and activate transcription from the PNfA3 promoter. As described (18), the phage ϕ29 *Hind*III H fragment produced two transcripts very efficiently: a protein p4-dependent one, 87 nucleotides (nt) long, originated at the PA3 promoter, and a protein p4-independent one, 470 nt long, originated at the PA2c promoter (Figure 4A and C). Transcription from the PA2b promoter was undetectable since it is not efficiently transcribed *in vitro* at the ionic strength conditions that are required for specific protein p4-dependent activity of the PA3 promoter (11). As shown in Figure 4A, the phage Nf *Hin*fI−*Sau*3A fragment was also efficiently transcribed from two promoters giving rise to a protein p4-dependent transcript, about 160 nt long, and a protein p4-independent one, about 410 nt long, whose sizes suggested that they could have been initiated at the PNfA2c and PNfA3 promoters, respectively (Figure 4C). Interestingly, phage ϕ29 protein p4 activated transcription from the putative late promoter of phage Nf suggesting that it had recognized the putative gpF binding site. No PNfA2b promoter transcription could be detected. To analyze if the ionic conditions were responsible of the PNfA2b promoter inactivity, as they were for the phage ϕ29 PA2b promoter, run-off assays were performed at lower ionic strength conditions in the absence of protein p4. As shown in Figure 4B, transcription from the phage ϕ29 PA2b promoter became detectable in these conditions, producing a 570 nt long transcript. When the phage Nf fragment was used, a ~510 nt long transcript appeared suggesting that the PNfA2b promoter was also active in these conditions (Figure 4C). Therefore, the three promoters detected in phage Nf behaved *in vitro* in a similar way to the PA2b, PA2c and PA3 promoters of phage ϕ29, suggesting that they might have equivalent roles *in vivo*. Nonetheless, no complementation *in vivo* was detected using *sus* mutants of phages Nf and ϕ29 (7).

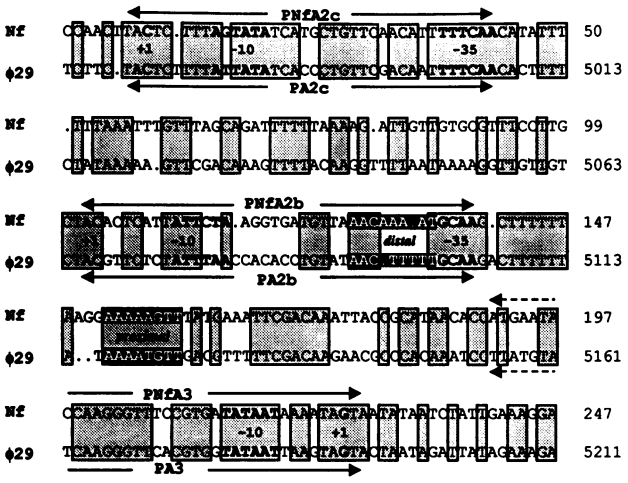

Figure 2. Sequence alignment of the region that controls transcription at late times of infection in bacteriophage ϕ29 (4964−5211) with the homologous region in bacteriophage Nf (1−249). Gaps are represented by dots and identical positions are enclosed in shaded boxes. The +1 position and the −10 and −35 consensus sequences of the promoters in both phages, are in bold. Promoters are indicated by horizontal arrows. The arrows indicating the upstream region of the PA3 and PNfA3 promoters are dashed since a −35 consensus sequence is missing. The protein p4 and the putative gpF binding sites are shown in white letters enclosed in black boxes, and their position relative to the late promoter is indicated in italics.

Binding of phage ϕ29 protein p4 to the putative gpF binding site in phage Nf

Since protein p4 was able to activate transcription from the phage Nf PNfA3 promoter, we analyzed whether it was binding to the putative gpF binding site despite its sequence divergence from the protein p4 binding site in phage ϕ29 (see above). As shown in Figure 5A and summarized in part B, protein p4 was indeed able to recognize the putative gpF binding site. The DNase I footprint of this DNA region, in the absence of any DNA binding protein, is characterized by cuts every 10 bp leaving the intermediate positions almost unaffected. These areas contain A- and T-tracts that have been described as poor substrates for DNase I (19). Binding of protein p4 was revealed by enhanced cleavage at positions −65 and −105, relative to the putative transcription initiation start site, and by the increased protection of the intervening areas. As it happens with the ϕ29 late promoter, RNA polymerase bound with low efficiency to the putative Nf

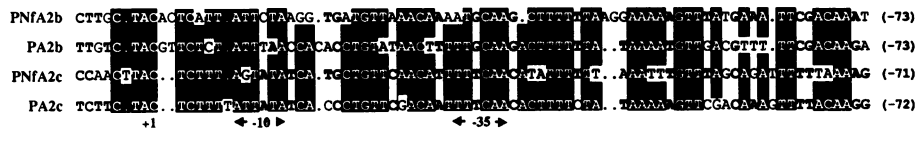

Figure 3. Sequence alignment of the phage Nf and phage ϕ29 early promoters. The transcription start site and the consensus −10 and −35 sequences are indicated. Conserved positions are in bold when present in two promoters, or enclosed in black boxes when present in three or four. Positions corresponding to the two protein p4 binding sequences in the PA2b promoter are indicated by lines above the PNfA2b sequence.

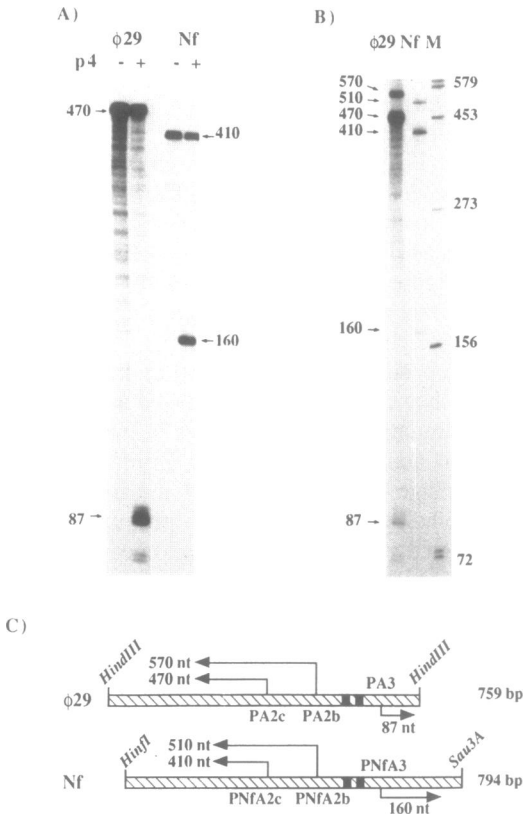

Figure 4. *In vitro* activity of the *Hind*III H and *Hinf*I−*Sau*3A DNA fragments from bacteriophages φ29 and Nf, respectively, in the absence or presence of protein p4 at high ionic strength conditions **(A)** and in the absence of protein p4 at low ionic strength conditions **(B)**. The origin of the fragment and the addition of protein p4 are indicated at the top of the run-off. Numbers on the side of the run-off indicate the size of the transcripts, if they are accompanied by arrows, or the size of the fragments used as molecular weight markers in lane M. **C)** Promoter arrangement in bacteriophage φ29 *Hind*III H fragment and that proposed in the *Hinf*I-*Sau*3A fragment of bacteriophage Nf. The protein p4 and the putative gpF binding sites are shown as black boxes. The size of the transcripts has been also indicated.

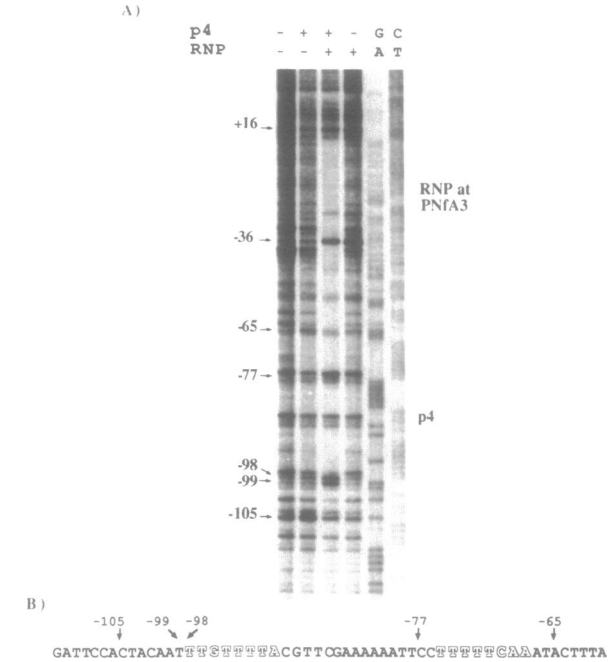

Figure 5. DNase I footprint of protein p4 and/or RNA polymerase bound to the putative late promoter of bacteriophage Nf containing the putative gpF binding site. **A)** The strand analyzed is the coding strand for early genes. Added proteins are indicated on the top of the footprint. Regions bound by the different proteins and positions that become hypersensitive to DNase I cutting are shown on the side. G/A and C/T are purine and pyrimidine sequencing ladders, respectively. **B)** Sequence of the putative gpF binding region. The sequence is that of the coding strand for early genes. Positions that become hypersensitive upon protein p4 binding are indicated with arrows and are refered to the transcription start site of the PNfA3 promoter. White letters indicate the putative protein gpF recognition sequences in phage Nf.

late promoter, inducing a hypersensitive site at position −36. In the presence of protein p4, RNA polymerase binding was much more efficient, as in the case of the φ29 promoter, and the PNfA3 promoter was protected down to position +16. Protein p4 binding was improved in the presence of the RNA polymerase, due to the existence of cooperativity between the two proteins (11), enhancing the protection pattern. Additionaly, some differences were detected in the protein p4-induced hypersensitivity pattern: the hypersensitive site at position −105 dissappeared whereas two new ones became apparent at positions −99 and −98. Similar differences are also produced in the protein p4 footprint when the RNA polymerase binds to the phage φ29 late promoter (B.Nuez, F.Rojo and M.Salas, unpublished), suggesting that the complexes formed by protein p4 and RNA polymerase are analogous in the two promoters.

DISCUSSION

The analysis of evolutionary related species frequently allows to evaluate the relevance of certain structures and/or mechanisms for the viability of different organisms. Critical mechanisms are likely to be the most conserved ones since any change leading to reduced efficiencies will disappear. Bacteriophages φ29 and Nf are evolutionary related (1, 2). They share a similar genetic organization (3, 4) and require the presence of an early viral protein for late transcription control (5, 6). In this report we provide evidence suggesting that the mechanism by which late transcription is controlled in phages Nf and φ29 is likely to be similar. In phage φ29, this mechanism involves binding of protein p4 to a region located between the early PA2b and the late PA3 promoters, activating transcription of late genes while inhibiting transcription of most of the early ones. We have identified a DNA region in phage Nf that contains putative promoters for the transcription of early and late genes, and a putative binding site for its transcriptional regulator, protein gpF. Moreover, the arrangement of promoters and transcriptional regulator binding site parallels that of the regulatory region for late transcription in phage φ29. The identified promoters were named PNfA2b, PNfA2c and PNfA3 since they correspond to the phage φ29 PA2b, PA2c and PA3 promoters, respectively. The *in vitro* activity of the phage Nf promoters correlated well with that of their phage φ29 homologues: the PNfA2c promoter was stronger than the PNfA2b promoter, being the latter salt-sensitive, and the PNfA3 promoter was transcribed efficiently only in the presence of protein p4. Therefore, the promoters identified in phage Nf could be equivalent to the phage φ29 promoters in transcription regulation at late times of the infection. The observation that protein p4 was able to bind to the putative protein gpF binding site and activate transcription from the PNfA3

promoter of phage Nf, strongly suggest that the transcriptional activation process of gpF and protein p4 could be analogous. Interestingly, protein p4 specifically recognized the putative gpF binding site, although it contains a spacer DNA 1 bp longer and an A-instead of a T-tract in the distal sequence. It has been described that protein p4 specifically recognizes the 5'-AA-CTTTTT-(15 bp)-AAAATGTT-3' inverted sequences (9; 10); however, change of the T- into an A-tract, as it occurs in phage Nf, seems to be a conservative substitution since it still provides protein p4 binding signals. In fact, recent experiments have revealed that protein p4 binds to the 5'-AACAAAAA-(15 bp)-TTTTTGTT-3' sequences (Nuez and Salas, submitted), in which the composition of the two original tracts has been inverted. A- and T-tracts are characterized by adopting a specific structure which is more rigid than B-DNA (20), therefore, protein p4, and maybe protein gpF, might recognize the particular structure of the tract rather than its base composition at their binding sites. Sequence alignment of the A2b and A2c early promoters contained in both phages has revealed the existence of internal homology between them, suggesting the duplication of a common unique promoter, that might have also contained an original transcriptional regulator binding site, prior to phage divergence. Such a duplication of the original promoter might have favoured a finely tuned transcriptional regulation providing a safety mechanism for the transcription of early genes.

ACKNOWLEDGMENTS

We are grateful to Dr L.Blanco for helpful discussions, to J.M.Lázaro for purification of protein p4 and RNA polymerase, and to L.Villar for technical assistance. This investigation has been aided by Research Grant 5R01 GM27242-13 from the National Institutes of Health, by Grant PB90-0091 from Dirección General de Investigación Científica y Técnica, and by an Institutional Grant from Fundación Ramón Areces. B.Nuez was holder of a pre-doctoral fellowship from Ministerio de Educación y Ciencia.

REFERENCES

1. Salas,M. (1991) *Annu. Rev. Biochem.*, **60**, 39–71.
2. Yoshikawa,H., Elder,J.H. and Ito,J. (1985) *Gene*, **37**, 125–130.
3. Mellado,R.P., Moreno,F., Viñuela,E., Salas,M., Reilly,B.E. and Anderson,D.L. (1976) *J. Virol.*, **19**, 595–500.
4. Matsumoto,K. and Hirokawa,H. (1981) *Mol. Gen. Genet.*, **184**, 180–182.
5. Sogo,J.M., Inciarte,M.R., Corral,J., Viñuela,E. and Salas,M. (1979) *J. Mol. Biol.*, **127**, 411–436.
6. Miyagi,T. (1985) Studies on Phage Structures Coded by Bacteriophage M2. M. Sc. Dissertation. Graduate School of Biological Science. Sophia University, Tokyo.
7. Mizukami,Y., Sekiya,T. and Hirokawa,H. (1986) *Gene*, **42**, 231–235.
8. Mellado,R.P., Barthelemy,I. and Salas,M. (1986) *J. Mol. Biol.*, **191**, 191–197.
9. Barthelemy,I. and Salas,M. (1989) *J. Mol. Biol.*, **208**, 225–232.
10. Rojo,F. Nuez,B., Mencía,M. and Salas,M. (1993) *Nucleic Acids Res.*, **21**, 935–940.
11. Nuez,B., Rojo,F., Barthelemy,I. and Salas,M. (1991) *Nucleic Acids Res.*, **19**, 2337–2342.
12. Rojo,F. and Salas,M. (1991) *The EMBO J.*, **10**, 3429–3438.
13. Barthelemy,I., Salas,M. and Mellado,R.P. (1986) *J. Virol.*, **60**, 874–879.
14. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor.
15. Inciarte,M.R., Lázaro,J.M., Salas,M. and Viñuela,E. (1976) *Virology*, **74**, 314–323.
16. Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
17. Rojo,F., Zaballos, A. and Salas,M. (1990) *J. Mol. Biol.*, **211**, 713–725.
18. Barthelemy,I., Lázaro,J.M., Méndez,E., Mellado,R.P. and Salas,M. (1987) *Nucleic Acids Res.*, **15**, 7781–7793.
19. Drew,H.R. and Travers,A.A. (1984) *Cell*, **37**, 491–502.
20. Levene,S.D., Wu,H-M. and Crothers,D.M. (1986) *Biochemistry*, **25**, 3988–3995.